NUFFIELD COUNCIL ON BIOETHICS

The collection, linking, use and exploitation of biological and health data: ethical issues

May 2014

Consultation Summary

Introduction and summary of process

A public consultation was held between October 2013 and January 2014 as part of the Nuffield Council on Bioethics' ongoing project on ethical issues in the collection, linking, and use of data in biological research and health care. A consultation document posing seven core questions and providing relevant background information was prepared by the Working Party. Respondents were encouraged to answer as many questions as they wished. The consultation document is available on the website of the Nuffield Council on Bioethics.

The consultation document was disseminated to a variety of individuals and organisations via a targeted mail-out as well as publication on the Council's website. 51 responses were received in total. 29 were from organisations and 22 were personal responses.

Respondents were also asked to indicate the nature of their interest in the consultation. The table below details the number of respondents for each interest. Please note that 26 of the respondents indicated more than one interest that applied to them/their organisation.

Interest	**Number of Respondents**
Academic	21
Regulatory/Legal	14
Biomedical Research	15
Other Professional Interest	14
Personal interest	6
NGO	6
Data protection officer	1
Information Technology professional	5
Knowledge and information management	5
Other general interest	4
Data owner	5
Government	4

Clinician	4
Professional interest - Caldicott guardian	1
Other - stated on form	5

1. Do biomedical data have special significance?

Is it useful (or even possible) to define biomedical data as a distinct class of data? If it is, what are the practical and ethical implications of different ways of defining this class?

'Biomedical data' was not generally regarded by respondents as a useful category for legal or regulatory purposes. It was pointed out that a number of different ways of grouping data exist.[1] Although these distinctions are not always straightforward, it might be sufficient to focus on the distinction between 'personal' and 'sensitive personal' data.[2] Nevertheless, it was suggested that a new category of 'biomedical data' might be linked to the Council of Europe's Convention on Human Rights and Biomedicine (which also refers to 'biomedicine' and 'biomedical research')[3] and – given the increasing amounts of behavioural data collected and used – renamed 'data on human health and behaviour' or 'human biomedical and behavioural data.'[4]

Concerning practical implications, several respondents recognised that the notional class of 'biomedical data' can cover a wide range of different forms of information/ data with different ethical sensitivities, varying according to context. Although overall the introduction of new regulatory categories did not seem to be favoured, individual or group level sensitivities can be complicated to assess within the few existing categories, to which increasing data sharing and linkage might contribute.[5]

What factors contribute to the belief that personal biomedical data deserve special protection? Does the sensitivity of biomedical data depend entirely on context or do biomedical data have special attributes that make them intrinsically more sensitive than other kinds of data?

This question is generally considered to be crucial. It did seem to create uncertainty, however, as a number of respondents interpret 'specialness of biomedical data' as well as the connotations and potential practical consequences of the distinction between 'intrinsic sensitivity' and 'context-dependent sensitivity' very differently.[6] A general potential for sensitivity, and that some data, e.g. data on mental or sexual health, are particularly sensitive, is a position that most respondents seem to accept. This is often linked to personal, identifying and predictive data being seen as intimately connected to

[1] Research uses, e.g., can be distinguished into biological data use (from tissue), health data (from medical records) and social data (see response of the Mason Institute for Medicine, Life Sciences and the Law, University of Edinburgh, p. 1); the new EU proposal for Data Protection Regulation distinguishes 'genetic data', 'biometric data' and 'data concerning health' (Atina Krajewska and Ruth Chadwick, Cardiff Law School, Cardiff University, p. 2). Cf. also PHG Foundation, p. 3-4).
[2] Information Commissioner's Office, p. 3-4; PHG Foundation, p. 3-4; The Mason Institute, p. 1-2; Medical Research Council (MRC), p. 14.
[3] Cf. Krajewska and Chadwick, p. 2.

[4] eHealth Research Group, University of Leeds Institute of Health Sciences, p. 1.
[5] E.g. response by the Information Commissioner's Office; cf. also responses to question 2.
[6] Cf. Prof. Tim Spector, KCL, p. 1; PHG Foundation, p. 3; McCormack, Woods and Leach-Scully, p. 1-2; Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, p. 1, and responses to question 2. One respondent considered the question "potentially misleading" (UCL Centre for Health Informatics and Multiprofessional Education, p. 1).

personal identity and a sphere of personal autonomy and privacy (concerning e.g. lifestyle) over time. It is also pointed out by some as rightly acknowledged by human rights law and current data protection regulations as (intrinsic) potential for sensitivity.[7]

While the use of terminology such as 'intrinsic sensitivity' might not necessarily be a helpful marker of risk, a number of respondents highlight the varying, context-dependent sensitivities and potential for harm. Indeed diverse social contexts or political systems can make differentiations in use and access to personal biomedical data acute. The potential to reveal predictive information on health risks or enable the identification of patients or research subjects is widely recognised to create the possibility of privacy breaches through misuse of personal data, stigmatisation and discrimination, in particular of vulnerable individuals and groups.[8]

Do some subsets of biomedical data (such as genomic data sets) present particular ethical challenges or offer ethically important benefits?

To what extent should genomic data sets be regarded as belonging to one individual and to what extent should other interests (e.g. of family members sharing genomic sequences) be recognised? What implications might this have for consent to collection of such data, for feedback concerning the data and for its broader use?

Context-dependent sensitivity of biomedical, including genomic data, is affected by the availability of new technologies. As many respondents point out, genomic datasets have an intrinsic potential for being sensitive, but none of the respondents took a strong 'genetic/genomic exceptionalist' position.[9] The sensitivity of biomedical/genomic data and risks linked to disclosure cuts across various potential categories of biomedical data. Although the full genome sequence provides a unique personal identifier, the same is true for other biometric data sets that require less specialist knowledge to use as identifiers.[10] The potential for identification might not be a good basis for regulatory measures, as some respondents conclude; other data (for example**,** social services data) are less exclusive than the DNA profile but might be more easily used to identify an individual and present greater risk if disclosed. In line with this, a number of people suggested that regulation should be proportionate to the risks presented by disclosure rather than linked to a specific data category.[11] In addition, given a variety of conditions, such as difficulties in quantifying risk and a more general tendency towards increased surveillance, precaution and attention to future developments in this area were felt to be necessary.[12]

Nevertheless, certain features of genome data were generally felt to require special consideration, such as their reliability and durability as a unique identifier, their links to family members, and its predictive value in relation to some health conditions (which can also be relevant for family members).[13] The potential to predict health outcomes, and the

[7] Cf. McCormack, Woods and Leach-Scully et al., p. 1; Sylwia Maria Olejarz, p. 1; Information Commissioner's Office, p. 1. This raises issues of defining personal identity and its uniqueness that are only hinted to in the responses. Also, not only personal, or rather, individual, identity is at stake. Both ordinary health data and genomic data are 'special' in that they relate to more than one data subject (Information Commissioner's Office, p. 4).

[8] "Biomedical data then seem to have no special significance in themselves, but do appear to have a very considerable significance in almost any social context – more perhaps in a rigidly controlled society", (Dr J. Saunders, Chair, Royal College of Physicians (RCP) Committee on Ethical Issues in Medicine, p. 1). Cf. also PHG Foundation, p. 4; McCormack, Woods and Leach-Scully, p. 1; The Mason Institute, p. 2; Farr Institute @ CIPHER, p. 1; The Wellcome Trust, p. 2; GeneWatch UK, p. 11.

[9] Cf. Krajewska and Chadwick, p. 3.

[10] Cf. Krajewska and Chadwick, p. 7; The Wellcome Trust, p. 2; Privacy Advisory Committee, Northern Ireland, p. 2.

[11] Anonymous respondent 1, p. 3; Information Commissioner's Office; cf. also answers to question 7.

[12] Cf. P. Finlay, p. 1-2; UCL Centre for Health Informatics and Multiprofessional Education, p. 2; GeneWatch UK; McCormack, Woods and Leach-Scully, p. 3.

[13] Cf., interestingly: "over time the "special position" of genomic information will progressively merge with the more general problem of providing access to all types of information that uniquely identify an individual. The ethical framework is likely to require a switch from trying to guarantee that the information does not uniquely identify a particular individual towards preventing inappropriate use of information by researchers, or to inadvertent use", Anonymous respondent 1, p. 1.

personal and familial impacts of this information, were widely perceived as an important ethical challenge in the consultation responses.

Concerning genomic data ownership a number of respondents tend to regard the individual source of genome data as a prima facie 'owner' of that data. However, given that these are not to be regarded as intrinsically more sensitive than other kinds of data sets, and the links to information concerning other individuals and groups, it is not clear what the implications of establishing rights of ownership would be.[14]

On the whole, the sensitivity and potential for harm of disclosing or misusing biomedical data is seen to be dependent on critical contexts and therefore governance measures should also be context-sensitive. In terms of consent, this would imply for several respondents that not all data uses will require explicit, written consent, since this could be disproportionate, and also unrealistic – the right to object to data use (like the right to privacy) is not absolute.[15]

2. What are the new privacy issues?

Do new information technologies and 'big data' science raise privacy issues that are new in kind or in scale?

Although responses illustrate some uncertainty about semantics, it seems generally uncontroversial that the scale of issues has changed because of new technological possibilities.[16] Some respondents argue or at least contemplate that this (might) entail qualitative changes. For example, as one respondent notes, the ambition of researchers, research funders and governments change in relation to the scale of data that can be analysed.[17] This includes multiple new interests in data as well as more possibilities for secondary use: using data provided not only for teaching, health (care) research, public health monitoring, disease registries and infectious disease reporting, but also for 'evidence-based health care practice' or a "rational approach to service provision; planning; financial management; commissioning of services; investigating complaints; auditing accounts."[18]

Advances in data science and technologies may have an impact on the overall quality and the nature of practice in research and health care because, for example, data are collected in anticipation that researchers or other people will make use of them in some way in the future.[19] This may lead to a different conceptualisation of 'privacy', and the social contexts in which it is at play. A change in temporality and dynamics in the system of norms surrounding 'privacy'[20] could be regarded as a change in kind. Both explicit and implicit presuppositions in ethics and governance of privacy might be undermined by these developments.[21] Several respondents mention that more and increasingly

[14] Cf. Ian Herbert, p. 15.
[15] E.g. PHG Foundation, p. 2; Krajewska and Chadwick, p. 16. A cultural trend of interpreting this increasingly 'permissive' is also observed that remains to be evaluated, cf. McCormack, Woods and Leach-Scully, p. 2.
[16] E.g. Prof. Neil Lawrence, Department of Computer Science and Sheffield Institute for Translational Neuroscience, University of Sheffield, p. 1; Farr Institute@ CIPHER, p. 2; MRC, p. 8-9. Cf. National Bioethics Commission of Mexico, p. 4.
[17] "Using NHS patient data for research is not new – what is new is the scale of the ambition. An ambition which will see an evolution from a cottage industry where researchers mined a specific set of data for a specific research project, into a Ford-like process where the collection and manipulation of these data is automatic, large scale, mechanised and slick", Association of Medical Research Charities (AMRC), p. 4.
[18] Privacy Advisory Committee, Northern Ireland, p. 1.
[19] This includes an increasing focus on "data-driven" science, cf. question 3.

[20] Cf. Krajewska and Chadwick, p. 5.

[21] Cf. "The DPA 1998 works best when data is collected for a well-defined and fairly stable set of purposes and users known beforehand. Its application to the creation and use of comprehensive databases of linked data collected/created prospectively for largely unknown purposes and users is much more problematic."; "Big Data enthusiasts seek to remove data protection mechanisms, as they have successfully done in the English HASC Act 2012", I. Herbert, p. 8; 14; "Widespread linkage risks blurring the distinction between research and service. This should not be done lightly, because the distinction has served us

competing interests are of ethical relevance, e.g. commercial interests in generating profits from data-mining medical records and other personal data, and evolving issues in relation to intellectual property protection and patient control over data,[22] government interest in public-private partnerships to use personal data to generate investment, income and economic growth; and government interest in gaining access to data for surveillance purposes.[23]

What are the implications for individual anonymity of linking data across large numbers of databases?

This was a main point of contention. A number of respondents note that anonymity cannot be guaranteed any longer, and that there is an increased potential for re-identification and deductive disclosure.[24] On the other hand, it is asserted that "Very few secondary purposes can only be achieved using identifying data, and linking data from different data sources is no exception."[25]

Even though a number of (primary) purposes might be served well with reliably anonymised data, it is not apparent that secondary uses are well-defined for the average data donor or the public. Many responses emphasise that there seems limited access to the relevant information.[26] That anonymisation of data may be possible in a variety of way and is (perhaps) generally reliable at present does not mean that this will remain the case in the future.

What is the 'public interest' in biomedical data? What benefits do we want to obtain? In what circumstances might the public interest take precedence over individual and minority group interests?

Anticipated benefits mentioned widely include improved prevention and treatment of ill health through more adequate and holistic research and more effective use of data for epidemiology and clinical trial recruitment.[27] Scientists and research funders, in particular, underline the substantial promise of personalised medicine – a better understanding of the interaction between genes and environment, and their impact on disease phenotype and epigenetics research, which could lead to better integration of lifestyle and health management for patients and the public generally.

Research and health care benefits are not clearly distinguished in this context, but are often referred to as apparent and generally in the public interest. Although there is also a general recognition that many potential public and individual health benefits may follow from an increase in private-public partnerships in health care and research using biomedical data, a number of respondents were very sceptical about 'public interest' being used rhetorically or strategically that might divert attention from more controversial data uses and about the longer-term scientific and political agenda-setting behind these, from which they feel excluded.[28]

Yet, some respondents also highlighted that there is a significant ignorance or under-appreciation of the moral harms and risks involved in not sharing data, including personal

---

well." (Progress Educational Trust, p.5); "the well-established principle that for consent to be meaningful it must be given in the context of a specific purpose, can be undermined with big data" (Information Commissioner's Office, p. 5); cf. also question 3.

[22] MRC, p. 7.

[23] Cf. Progress Educational Trust, p. 5-6; GeneWatch UK.

[24] E.g. Information Commissioner's Office, p. 4; McCormack, Woods and Leach-Scully, p. 3; Prof. Sheila M. Bird; Dr J. Saunders, p. 3.

[25] I. Herbert, p. 7.

[26] Cf. question 7.

[27] E.g. The Physiological Society; MRC; British Dental Association.

[28] Cf. in particular GeneWatch UK, and responses to question 7.

biomedical information.[29] Others think that this argument is used as a 'Trojan horse' to undermine important human/civil rights or, at least, that arguments using appeal to 'public interest' are not as strong in the scenarios that will develop in biomedical data use as in traditional public health cases such as infection control.[30]

A few responses suggest that there is no inherent dilemma between 'privacy' and 'public interest', and both are seen as interrelated, 'moving targets'.[31]

What are the actual harms we should seek to avoid in using biomedical data (e.g. discrimination, stigmatisation)? What evidence is there of these harms having occurred?

Most responses converge on potentially problematic issues such as stigmatisation, discrimination, coercion, exploitation, misrepresentation, abuse of vulnerable parties such as political dissidents, abused women and children, or women who have hidden the paternity of a child. Concerns were expressed about safeguards against abuse of technologies by authorities or criminals, and about the security, privacy and integrity of personal information. Yet, the responses generally reflect a striking disagreement on the imminence and/or severity of risks, with responses ranging from "most fears [are] irrational and risks are never quantified"[32] to "there is real harm we need to seek to avoid."[33] Some responses call for the need to establish more evidence through research.[34]

In what ways does it matter if people's data are used in ways of which they are unaware but that will never affect them?

A number of respondents relate this question to the requirement for informed consent and point out that it is an important signifier of respect for an individual's autonomy as well as of trustworthy relationships with doctors, researchers and the health care system, and indeed that it would be 'disturbing' if this were considered open to debate. A few respondents associated this question with other issues concerning the interpretation of valid consent requirements in new research and data use contexts, such as the potential of consent to adapt to wider secondary and future uses, but also questions concerning the moral foundations of consent requirements, e.g. what kind of harm it should seek to prevent, on redefinitions of autonomy and ownership of biomaterial and data, and the need to specify public benefits of big data research.[35]

---

[29] E.g. P. Finlay, p. 4; Anonymous 1, p. 8; Cancer Research UK; Wellcome Trust Sanger Institute.
[30] Cf. PHG Foundation, p. 6.

[31] Cf. Krajewska and Chadwick, p. 17-18.
[32] Prof. T. Spector, p. 2.

[33] WMA, p. 2. Cf. "recent research (triangulated approaches and validation studies) conducted in Australia, Canada, the US, and Europe seems to suggest that genetic discrimination is now an established, incontrovertible ethical, legal, and psychosocial phenomenon" (Krajewska and Chadwick, p. 10).

[34] McCormack, Woods and Leach-Scully, p. 3; The Wellcome Trust, p. 4. A noteworthy proposal is the setting up of an "open anonymous web based register of confirmed harms arising from privacy breaches to which the public and/or professionals would be encouraged to contribute" eHealth Research Group, University of Leeds Institute of Health Sciences, p. 1.

[35] "It matters because it affects their autonomy and the basic principle that they need to decide what could be done with their information. Furthermore, the individuals have to decide by themselves what can affect them or not, especially take into account that these data could be used in the future." (National Bioethics Committee Mexico, p 5); "people are keen to take part in medical research, but only when they have been asked. This is an important safeguard to protect not only individual privacy but the broader public interest. (reference to Wellcome Trust-commissioned research, GeneWatch UK, p. 3); "Asking for consent to before using identifiable data for secondary purposes is seen as a mark of respect"; "Please explain what is meant by 'will never affect them'. Just becoming aware that this is generally happening would seriously erode the trust that is the basis of the patient-clinician relationship, which is definitely not in the public interest (vide the emphasis on "no surprises" in the 2003 NHS Code of Practice Patient Data Confidentiality). I find it disturbing that it should be considered necessary to ask this." (I. Herbert, p. 6; 15).

How are applications of computer-based technology (e.g. social networking, image sharing, etc.) affecting concepts of privacy, identity and social relatedness? How are related behavioural norms influenced (e.g. willingness to share and publish data)?

Respondents generally seemed to agree that the public is behaving ambivalently:[36] on the one hand more personal data sharing can be observed when users see this as beneficial for themselves;[37] on the other, data sharing generates concerns, albeit ones that remain relatively diffuse.[38] Some responses call for caution concerning regulatory and/or ethical conclusions from the observation of people's behaviour.[39]

Would it be helpful to treat biomedical data as 'property'?

Roughly equal fractions of responses are in favour of applying property terms and constructs in this context, are against this, or consider it more neutrally a topic for potentially relevant further investigation.[40]

Arguments in favour in some responses assume that the current or default situation is or should be interpreted as one of individual data 'ownership', implying that biomedical data use and access are limited by the owners' consent. This is seen to strengthen the position of the data source, including, in particular, a right to receive feedback on relevant research outcomes.[41]

Responses that do not regard property as a useful reference point out that property in data is not meaningful in a legal sense, and that governance should rather focus on benefits and burdens of data access and use.[42] Individual data ownership might hinder effective research and wider health data use without, however, having clear value for the individual.[43]

3. What is the impact of developments in data science and information technology?

To what extent and in what ways has the availability of biomedical data and new techniques for analysing them affected the way in which biomedical research is designed

---

[36] E.g. McCormack, Woods and Leach-Scully p. 3-4; The Mason Institute, p. 8.

[37] E.g. self-help e-communities. Farr Institute@CIPHER, p. 3.

[38] "It is clear that computer-based technology applications are affecting the concepts of privacy, identity and social relatedness. For instance, knowing the people that we are interacting with is no longer necessary, the veracity of the information shared cannot be trusted, and people are becoming prone to interact with machines instead of people" (National Bioethics Commission of Mexico p. 5); There could be a "distillation of effects" through social networking and commercial data collection. The ability to link, for example, loyalty card data with genomic data, gives a relatively complete picture of the environmental and biological background for an individual. Holders of this data may be able to make relatively accurate predictions for future life outcomes for this individual." (Prof. N. Lawrence, p. 1).

[39] WMA, p. 2.

[40] Cf. "The implications of treating biomedical data as property deserve a detailed discussion. Such a status is at odds with (non) ownership of bodies and concepts of 'stewardship", Dr J. Saunders, p. 4; "it is possible to acknowledge a database as property but one that is held in trust by the custodians of the data" (McCormack, Woods and Leach-Scully, p. 3), implying differing views on stewardship/ trusts arrangements.

[41] Cf. National Bioethics Commission of Mexico, p. 5; Information Commissioner's Office, p. 6-7; S. Olejarz, p. 1; this is also acknowledged by responses overall not in favour of treating biomedical data as "property: "A property-type paradigm is likely to give rise to expectations that feedback is part of the core set of entitlements arising from uses of (biomedical) data" (The Mason Institute, p. 5.

[42] The Mason Institute, p. 4-5.

[43] "Ownership is a barrier to making data openly available, particularly when the data have been collected by consortia or consists of an aggregation of data from different sources." Royal Academy of Engineering, p. 5; "If these data are deemed to be "property" we suspect that distribution for research and public health uses will be curtailed, or made more difficult, while having little additional value for the individual", Anonymous response 1, p. 4.

and funded? Is there any evidence that these factors have affected (or are likely to affect) research priorities?

Respondents agree that the technical capacity to gather large amounts of data generates interest in more and widely differing uses: these include, in addition to extensive use for diagnosis and research, service improvement, both in health care systems and also multi-national companies (such as e.g. Google, Facebook and Amazon that engage in extensive 'harvesting' of big data).[44] In biomedical research, more specifically, there are considerable expectations of more systematic and precise or personalised medicine.[45]

Effects on research practice are described by several respondents: some consider that "ideological change is taking place in science which tends to abandon the idea of hypothesis-driven science and replaces it with mindless data mining"[46] and a number of respondents are sceptical or even extremely doubtful that the general approach or underlying focus on big data and genomics is a good research and health care priority.[47] It is argued in this context that data-driven research and a market-driven (instead of need-driven) approach to health care overall is undermining ethical and governance principles for research ('function creep'[48] and blurring of the distinction between research and service provision).

Commercial involvement is suggested to increase conflicts of interest between health care providers, and may limit data access, even in theoretically open environments (research data held by companies may not be shared).[49] It was argued that there should be commensurate investment in statistics and interdisciplinary research to interpret the data generated, and that we should not neglect alternative and (what some argued were) more cost-effective public health measures.[50]

Some respondents suggested that more attention should be paid to particular social groups for which, for example, genetic testing would be a health care priority. They argued that investments in data and tissue collection, data curation, infrastructure and tools, should be 'science-driven' – i.e. "tightly linked to well-developed visions for specific future scientific needs."[51] The big data approach can even lead to bad policy and more bias.[52] It was suggested that there is a need to be wary of overinflated promises for health impact, the political drivers and unintended consequences, and effects such as over-treatment and medicalisation, genetic determinism and equality of access to benefits.[53] Interestingly, these effects are explained also as a consequence of poor analytical and in particular mathematical/statistical skills.[54]

What are the main interests and incentives driving advances in data science and technology that can be applied to biomedical data? What are the main barriers to development and innovation?

Opportunities of big data in addition to the ones mentioned are in particular data availability and completeness.[55] However, there seems to be a perception that the challenges of defining or bringing together different types of data are substantial, being not only technical, but also social and epistemological. It was argued that interdisciplinary

[44] Farr Institute@CIPHER, p. 4.

[45] "Though studies recognizing the highly individualized nature of disease are becoming increasingly prominent—through, for example, "n=1" clinical trials (van der Greef et al., 2006) or studies of the "patient journey" (Kinross et al., 2011)—the outcome of personalized medicine is still predominantly about statistical likelihood, chance, and variance." (N. Levin, p. 3).
[46] GeneWatch UK, p. 13.
[47] UCL Centre for Health Informatics and Multiprofessional Education, p. 1-2.
[48] Royal Academy of Engineering, p. 4.
[49] Prof. N. Lawrence, p. 2; Krajewska and Chadwick, p. 19.
[50] Prof. C. Brayne, p. 2; 4; GeneWatch UK, p. 18.
[51] MRC, p. 2.
[52] Royal Academy of Engineering, p. 4.
[53] GeneWatch UK, p. 14; Prof. C. Brayne; cf. Martin Bobrow, p. 1.
[54] Prof. C. Brayne.
[55] UCL Health Informatics, p. 3.

training is needed to make big data approaches fruitful.[56] Further barriers mentioned include the protection of professional interests and of intellectual property, which can prevent data sharing, and also negative effects of commercialization more generally.[57]

One respondent commented that the language employed in this consultation question can be interpreted as tendentious: "It is not at all clear that the drives to create vast data banks are being done with population health or sustainability in mind. (...) The very language being used 'barriers to development and innovation' suggests a rush to change many things without proper evaluation".[58]

Does 'big data' need a more precise definition or is it a useful concept in the life sciences even if loosely defined? Has enthusiasm for 'big data' led to over-inflated expectations on the part of governments, researchers and/or the general public?

Although there is some imprecision and clearly inflationary use and expectations at this point, some respondents point out that also unique advances can be expected from 'big data research'.[59]

What are the significant developments in the linking or use of biomedical data, including any we have not mentioned, to which we should pay attention in our deliberations?

Issues mentioned include particular technological developments such as text mining and virtual research environments.[60] More generally, it was suggested, there should be more attention given to the issue of trust in expertise.[61] Several respondents also underline the importance of professional (ethical) standards[62] as well as the increasing integration of health and social care data.[63]

4. What are the opportunities for, and the impacts of, use of linked biomedical data in research?

What are the hopes and expectations associated with data use for biomedical, public health and life sciences research? What are the main concerns and fears?

Benefits of data linkage mentioned include, in particular, a more holistic view of certain medical conditions, and a better and faster way to diagnose, treat, and prevent disease. Although these techniques still need to be validated before routine clinical use, they might transform the concept of genetic testing and constitute a significant step towards more personalised medicine. Also, pharmaco-epidemiological studies might generate information about the frequency, effectiveness and adverse reactions to medicines. According to at least one respondent, on the other hand, "the evidence has grown that

[56] Cf. N. Levin, p. 1; I. Herbert, p. 16-17.

[57] "The main barrier are the great opportunities for economic gain that implies, these creates an environment where development and innovation is carried out exclusively if an economical profit is foreseen, and excludes or limits other areas of development" (National Bioethics Commission of Mexico, p. 6).

[58] Prof. C. Brayne, p. 2.

[59] "The experience shared by single researchers, large organisations, and everyone in between is that at every scale our ability to generate data is growing faster than our ability to manage, store, and analyse those data. This, perhaps, is how "big data" should be defined: not as a quantity, but as a rate." Anonymous respondent 1, p. 5; cf. M. Bobrow, p. 1.

[60] eHealth Research Group, University of Leeds Institute of Health Sciences, p. 2.

[61] UK GeneWatch p. 3.

[62] Exeter Centre for the Study of the Life Sciences (Egenis), p. 2; AMRC, p. 8; MRC, p. 13.

[63] "We recommend that consideration is given to the inclusion of social and social care information within this consultation. In Northern Ireland there has been an integrated approach to health and social care for many years. This is the direction of travel for the rest of the UK." Privacy Advisory Committee, Northern Ireland, p. 4.

genomic testing has poor predictive value for most diseases in most people and is unsuitable for use in newborn screening programmes."[64]

To what extent do the kinds of collaborations required for data-driven research (e.g. international or multi-centre collaborations) generate new ethical and social issues and questions to those in other forms of research?

Issues mentioned include the inapplicability of traditional consent standards and lack of democratic debate around big data projects.[65] In addition, intercultural issues tend to be disregarded, such as differing conceptions of privacy.[66] The involvement of more stakeholders was generally felt to be necessary.[67] There was some suggestion that it was also becoming increasingly difficult for researchers to understand their complex obligations.[68]

Should researchers be required to allow others to access data they have collected for further research?

Relatively few responses directly engaged with this question and the ones that do tend to reflect disagreement or opinions in transition. Some respondents mentioned that the move towards 'open science' clashes with the current culture and incentive system of scientific research.[69] Sharing is favoured by some responses on conditions of time-limited, privileged access and if funded publicly,[70] but also opposed, notably by the WMA.[71]

What sorts of concerns are raised when research is carried out by a commercial firm?

A number of responses note that the profit motive can undermine scientific integrity and potentially socially beneficial research and/or that differing 'commercial ethics' is in tension with people's trust in medical research.[72] This can lead to fears of exploitation of data sources and that commercial use can only insufficiently be controlled. In particular, minority interests in controversial research use might not be respected, data generally or negative research results in particular might not be shared, competition might dis-incentivise standardisation of data sets, and research may be obstructed by professional codes and freedoms.[73] An additional issue noted is cost-effectiveness for the public/the health care system users.[74] A related concern was that exclusivity of access to data can drive up prices for drugs. Overall, issues in this area seem to be considered as very important in ethical terms and also as a 'major regulatory challenge.'[75]

---

[64] GeneWatch UK, p. 1. For other, partly related, concerns and fears cf. question 3.
[65] E.g. Prof. T. Spector, p. 2; WMA, p. 5. Cf. general governance inconsistencies, question 7.
[66] McCormack, Woods and Leach-Scully, p. 4.
[67] PHG Foundation; the Mason Institute; Ian Herbert.

[68] Cf. McCormack, Woods and Leach-Scully, p. 4; WMA, p. 2-3.

[69] Cf. Krajewska and Chadwick, p. 14; N. Levin, p. 6; Dr J. Saunders, p. 5.
[70] Prof. T. Spector, p. 2; eHealth Research Group, University of Leeds, p. 3.

[71] WMA, p. 3.

[72] E.g. WMA; Nowgen consultation.
[73] Cf. problems with genetic testing companies, GeneWatch UK; Prof. C. Brayne.
[74] Prof. N. Lawrence, p. 2.

[75] PHG Foundation, p. 9; Prof. N. Lawrence. p. 2.

5. What are the opportunities for, and the impacts of, data linking in medical practice?

What are the main hopes and expectations for medical practice associated with increased use of linked electronic data? What are the main concerns or fears?

Opportunities pointed to by a number of respondents are improved care, improved monitoring of care, better patient outcomes, as well a better and earlier diagnosis of conditions (e.g. dementia) and potentially enormous savings in health care, although these would depend on substantial investment.[76]

Concerns mentioned are similar to the ones expressed with reference to commercialisation in research and in relation to big data approaches as the main research priority, although they are perhaps even stronger in this context. Indeed the impact on the public health care system is criticised very strongly by a few responses,[77] and as leading to mistrust in the medical profession, scientists and government more generally. The doctor-patient relationship, in particular, could be 'irreparably damaged' if patients receive insufficient information about data sharing.[78] Another noteworthy concern appearing in a few responses is that there is an emerging issue of liability if available data are not used for patient care.[79]

What can be said about public expectations about the use of health care data, in terms of appropriate use, information and control? To what extent would members of the public expect health care data to be shared with other agencies or bodies?

A few, general comments refer to the need for adequate balancing of increased data use and respecting people's wishes, and not evading controversial issues by relying on consent that is too broad or invalid. There was concern that the high level of trust that the public is generally assumed to have towards science and doctors, in particular, should not be misused. More specifically, it was highlighted by some that the public would expect data sharing to progress, but for valuable and clearly communicated purposes rather than the generation of profit.[80]

Is there potential for privacy controls to hide secrets, such as abuse, or to disadvantage people in unintended ways (by preventing best treatment, perhaps)?

While there are emerging liabilities if available data are not used to offer best treatment, most respondents would consider that particularly sensitive personal information, such as information concerning abuse or psychiatric treatment of individuals, should remain protected. In addition, there is also a danger that particularly vulnerable people might not seek treatment if they have to fear re-identification and stigma.[81]

Are there particular issues raised by 'risk-profiling' where individuals at high-risk (e.g. of type 2 diabetes) are identified and approached for specific interventions? What might make the difference between this being intrusive and it being supportive?

---

[76] MRC; The Wellcome Trust; N. Lawrence, p. 1; AMRC, p. 8-9; I. Herbert, p. 16; Royal Academy of Engineering, p. 5.
[77] GeneWatch UK.
[78] WMA, p. 3.
[79] Royal Academy of Engineering, p. 6.

[80] Cf. WMA, p. 2-3; Prof. C. Brayne; The Wellcome Trust, p. 8-9; GeneWatch UK; National Bioethics Commission of Mexico.

[81] I. Herbert, p. 6; GeneWatch UK. Cf. Prof. S. Bird, p. 2.

Among the relatively few comments in relation to this question the main issues mentioned were that the boundaries between health and ill-health can be unclear;[82] that the understanding of the concept of risk is often poor;[83] and that the availability of benefits and treatments should be considered if patients are identified as being at 'high-risk'.[84] More generally and with reference to future developments, it is pointed out that patient interest, rather than commercial interest, should be primary.[85] Risk communication is perceived as a problematic issue as well as who has access to such information.[86] Dangers of risk profiling are evaluated by some as substantial, including the potential to define an "unemployable and uninsurable" category of patient."[87]

What are the implications of episodes of treatment across different care providers being used routinely as research data? How might this affect the ethical basis of the doctor-patient relationship?

According to some, this could lead to improvement of disease diagnostics and care, although in the first instance, as a few responses emphasise, there is the "need to change the mindset of public and doctors that all data should be used for research."[88] Again, however, patients should be informed explicitly of what happens to the data in the interest of preserving trust.[89]

To what extent does the possibility that biomedical data can contribute to a research base to advance the effective treatment of others create a moral obligation to allow them to be used in this way? What might limit this obligation? How should we regard (and provide for) those who refuse to allow their data to be used?

There was general agreement among respondents that there is no (strict) moral duty to share (in particular identifiable) data, and that participation should be on the basis of mutual agreement, although a few respondents emphasise the advantages and benefits of data sharing and/or a general desirability that people participate in research or contribute to the public health care system.[90] However, most seem to think that participation should remain voluntary and based on 'mutual acceptance'. It is also suggested that differing contexts of sharing should be distinguished.[91]

The right to opt-out of sharing or research participation without any detrimental consequences for personal care is considered essential by several respondents[92] and threats to make care provision conditional as undermining consent as well as a 'bullying tactic'.[93]

[82] PHG Foundation, p. 12.

[83] Dr John Saunders, p. 5.
[84] I. Herbert, p. 21.
[85] S. Bird, p. 4.
[86] E.g. Nowgen consultation.
[87] PHG Foundation, p.9.
[88] T. Spector, p. 3; UCL Women's Health, p. 2.
[89] WMA, p. 2.
[90] Cf. Farr Institute@CIPHER, p. 4-5.
[91] E.g. "sharing in a commercial context does not involve any moral obligation, and sharing data for research should be distinguished from sharing for service development, e.g. clinical audit, financial management" (PHG Foundation, p. 12).
[92] WMA, p. 3.
[93] Prof. S. Bird, p. 3.

6. What are the opportunities for, and the impacts of, using biomedical data outside biomedical research and health care?

What are the main hopes and expectations associated with the wider use of biomedical data (outside biomedical research and clinical practice)? What are the main concerns or fears?

The main applications identified by respondents were in health and safety contexts, and for commercial purposes and/or economic development.[94] Data uses explicitly mentioned include consumer goods design (e.g. ergonomics) and drug development as well as lifestyle counselling and marketing.[95] These uses, however, create concerns that potential benefits will primarily accrue to businesses or to more affluent sections of the population, or overlap in uncontrollable ways with data uses that the data subjects would not anticipate or find controversial, such as 'bio-surveillance', which might affect, in particular, people vulnerable to stigmatisation and discrimination.[96]

What factors are relevant to determining the legitimate scope of further uses of biomedical data? For example, should it be restricted to a 'compatible purpose' (and, if so, how might this be defined)? To use only by public authorities (and those providing public services under contract)? To non-commercial or non-profit uses/ users?

Legitimacy is considered by many to be determined in the first instance by the individual data subject giving consent, yet, given that public and private interest in secondary uses of data overlap, some consider it to be simplistic and impossible to pre-define compatible purposes and/or suggest that the attempt to do so might lead to excessive bureaucracy.[97] Compatible uses should serve 'public interest' and "not line the pockets of individuals or organizations."[98] It is also more explicitly suggested in one response that, for example, industry-funded research should not be considered as categorically different from publicly (government or charity) funded research, under the precondition that the research has been approved by the appropriate scientific standards and passed ethics approval.[99] Others note that wider uses and sharing should take into account that the scientific value of data (for example, in risk prediction) can be limited and that market forces might dictate how the data are used.[100]

What are the ethical implications of using predictive analytic tools with biomedical data outside health care and research (e.g. in recruitment or workforce management)?

Predictive analytic techniques are likely to have data protection implications (compliance with fairness and proportionality requirements of DPA).[101] Concerns were expressed that the complex nature of IT systems might make it easy to hide the use of 'risk profiling' techniques for hiring decisions etc.[102] The current moratorium concerning the use of predictive genetic testing results for insurance purposes should take into account the rapid developments in the field.[103]

[94] Cf., however, the emphasis of 'biosocial' research opportunities (MRC).
[95] Royal Academy of Engineering, p. 6.
[96] Cf. questions 2. and 3.
[97] PHG Foundation, p. 13.
[98] WMA, p. 4.
[99] Cancer Research UK, p. 3; cf. The Mason Institute, p. 14.
[100] PHG Foundation, p. 13.
[101] Cf. Information Commissioner's Office, p. 9-10.
[102] eHealth Research Group, University of Leeds Institute of Health Sciences, p. 4.
[103] PHG Foundation, p.6; The Wellcome Trust, p. 10; cf. responses to questions 2. and 5. (new privacy risks and risk-profiling).

Would the ability of individuals to maintain direct control over the use of data about them be likely to affect the range of further uses to which they would allow the data to be put?

Some respondents suggested that the greater the control individuals had over the use of data, the greater the range of further uses might be possible – assuming that control is "real and that consent for use is truly fully informed."[104] On the other hand, there were concerns about the extreme difficulty and consequences of effectively removing data once they had been used in aggregate analysis.[105]

Should individuals be able to profit from the use of their biomedical data (e.g. by selling access to the data to commercial companies)?

Responses to this question echoed the ones to the earlier question about 'property' rights in data. Considerable concern was expressed about the implications of allowing individuals to sell and profit from use of 'their' data. Some respondents thought that although individuals should 'own' biomedical information about them they should not be able to profit from selling such data because this might involve coercion of vulnerable parties,[106] while others consider the situation to be analogous to banking services, as a form of licensing data for commercial use.[107] It was also pointed out that the distinction between commercial and non-commercial use of body parts and data is already blurred, and it should be made clearer why institutions should be able to profit from data sale while individuals can or should not.[108] Although it was thought to be difficult to ban or outlaw this development, a more widespread encouragement or policy of individual data selling was frowned upon by some as a motive for participating in research, which would contradict a widely-supported ethos of altruistic research participation.[109]

7. What legal and governance mechanisms might support the ethical linking and use of biomedical data?

What ethical principles should inform the governance of biomedical data? For example, should the principle of 'respect for persons' be given primacy here? How might this relate to principles such as solidarity and tolerance?

A variety of principles and moral values and concepts – autonomy, privacy, respect for persons, dignity (and the subject's 'property right') expressed by informed consent – were mentioned. Among these 'respect for persons' was seen by many as important, although not necessarily primary. Others put more emphasis on legal balancing principles of proportionality, necessity and risk assessment and/or criticised an overly strong focus on individual rights. Limits to the primacy of individual control were generally accepted, though considered as exceptional, that is, justified only or mainly by certain public health needs or in emergency situations. For some, however, these might include clearly defined public benefits of increasing data sharing.

Other noteworthy 'principles' mentioned were a "principle of deliberative democracy"[110] – the responsibility to engage the public on these matters; "non-discrimination and social justice (inclusion)," a "principle of transparency and openness"[111] as well as "good

[104] WMA, p. 4.

[105] Anonymous response 1, p. 9.
[106] Farr Institute@ CIPHER, p. 6.
[107] Prof. N. Lawrence, p. 2.
[108] PHG Foundation, p. 14; Prof. C. Brayne, p. 3.
[109] Anonymous response 1, p. 9.
[110] Cf. PHG Foundation, p. 16.
[111] Krajewska and Chadwick, p. 19.

citizenry with rights and obligations."[112] A recurrent theme was the interrelatedness between privacy protection and solidarity, or use of data for the public good.

Does the use of linked biomedical data require distinctive governance arrangements compared to the use of other personal data?

Respondents suggested distinctive requirements since established forms of ethics and governance were being stretched 'unsustainably' by current big data developments and by the fragmentation of research and health care systems,[113] including, in particular, the interface with the commercial sector. Governance arrangements were thought to require multidisciplinary expertise.[114] A context-sensitive approach was generally favoured, which should avoid being primarily prohibitive. Some respondents emphasised the need to recognise also a fragmentation of data uses and the importance of the 'data environment'.[115] A minority point of view was that biomedical data governance could be analogous to financial data governance.[116]

Are the current principles of consent – including the principle that consent can be withdrawn – still 'fit for purpose' in relation to the linking of biomedical data?

Many respondents identified informed consent as the main ethical and legal mechanism at stake, while noting a number of limitations to applicability of standard consent approaches for big data and data linking initiatives. Some suggest that these latter developments undermine the possibility of consent and that new data initiatives could lead to a violation of proportionality requirements and/or human rights (data retention for unspecified secondary purposes; problem of withdrawal of data from aggregate analysis; unclear risks; absent or insufficient risk-benefit analysis; no recognition of ongoing relationships).

Although stratification of consent was proposed as a viable solution,[117] many considered the use of consent mechanisms to be unfit to safeguard privacy rights. There were also concerns that stricter and more explicit consent requirements would not be in the data subject's or researcher's best interests.[118] On the other hand, a generic move to broad consent was considered problematic. In addition, there were concerns about insufficient communication of issues that will become increasingly relevant, such as the cooperation with commercial partners.

What level of continuing involvement is it reasonable to expect individuals to have in how their data are used after they have been collected?

Continuing involvement or information is considered important by a number of respondents, though there was some debate about how this could best be achieved. It was suggested that a minimal requirement would be an independent and trustworthy governance mechanism for the relevant project in which all data donors/research participants are represented and the data uses are assessed.[119] Some respondents argued that data donors have a right to feedback concerning research findings, and their wishes concerning ongoing involvement should be respected, which was recognised as an increasingly important challenge.

[112] Farr Institute@CIPHER, p. 5.
[113] McCormack, Woods and Leach-Scully, p. 1-2.
[114] GeneWatch UK, p. 4; National Bioethics Commission of Mexico, p. 10.
[115] Prof. C. Brayne, p. 3; The Mason Institute, p. 6. Cf. Prof. N. Lawrence, p. 2 on "data mutuals".
[116] Prof. T. Spector, p. 1; 3.
[117] AMRC, p. 1.
[118] Cf. The Mason Institute; MRC, p. 14-15.
[119] I. Herbert, p. 26.

Should there be an opt-in or an opt-out system for people to decide whether to allow their personal medical data to be used for public benefit?

This issue proved controversial, both on principled grounds and in more practical, governance-oriented terms. While some respondents insisted on "fully informed 'opt-in'",[120] others considered that there was a general moral duty to contribute to research and that specific consent is often disproportionate and/or impractical or unworkable.[121] However, it was widely recognised that opt-out systems can be prone to misuse, and should not result in the exploitation of people's ignorance; they should therefore be accompanied by public engagement campaigns.[122] Specific points made about NHS England's care.data programme included that the possibility of opting out should be made a statutory right.[123]

There was a lack of clarity about the meaning and scope of 'opt-in' and 'opt-out' systems, and a number of nuances were used ('effective opt-out', 'true consent').[124] Independently of preferences for either of such systems, there was a view that the concerns of data subjects need to be addressed.[125]

Under what conditions ought individuals to be content to delegate authorisation of the use of health and biological data about them?

Very few responses addressed this point. Among the most important issues mentioned by respondents were that governance approaches in general should be risk-based and adaptable, since it is not possible to outline specific conditions for authorisation in advance.

What role should public engagement and democratic processes play in the determination of governance measures? In what circumstances, if any, might the outcome of democratic procedures mandate overriding individual interests?

This was considered a main area of controversy and many respondents emphasised the lack of transparency and information available to the public about access and use of data, about the concrete benefits from data sharing and linkage, as well as the appropriate balance between privacy and 'public good'. A number of respondents expressed a suspicion concerning political influences and/or conflicts of interest more generally: "Public engagement and democratic processes have critical roles to play in the determination of governance measures. Many governments have an inherent conflict of interest in this area as they intent to use the data generated for their own planning purposes, and private companies seek to use it to generate profit."[126] Concerns were expressed that certain sections of the population would remain excluded from decisions.[127]

[120] GeneWatch UK, p. 20.
[121] The Wellcome Trust, p. 13.
[122] Cf. WMA, p. 4-5.
[123] AMRC, p. 9.
[124] E.g. British Dental Association, p. 2.

[125] "Whichever type of consent is sought, it ought to be clear to the participant how their data can be used, including: whether it would be accessible to commercial organisations; whether it will be possible to withdraw consent once data have been made available for access to researchers; and whether or not they will be informed of any health-related findings resulting from the processing of their data." (The Wellcome Trust, p. 11).
[126] WMA, p. 5.
[127] "there is also a major need for improved cultural awareness to ensure that sections of the population who are traditionally harder to reach – certain ethnicities, socioeconomic groups, and (in some cases) people affected by illness or disability – are adequately represented, informed, and consulted." McCormack, Woods and Leach-Scully, p. 5.

What inconsistencies exist in current ethical guidance and governance structures relating to biomedical data?

Many responses drew attention to an over-reliance on consent mechanisms, and there was broad agreement with the need to move beyond reliance solely on consent or anonymisation of data. A related proposal was that research should proceed by 'public consent' rather than with focus on the individual (project), although this would imply that research methods have to be justified to the public, otherwise this approach would risk supporting poorly designed research.[128] Problems of re-defining and adapting consent reappear in this context. Making consent and governance arrangements dynamic and flexible is also seen as a necessary consequence of the difficulties in informing data subjects of potential future uses.

Responses also drew attention to inconsistencies produced by overlapping legislation in the UK and internationally. Researchers sometimes appear to be unsure which rules and regulations to follow, and what their concrete obligations are. This situation was expected to worsen and to unnecessarily risk causing harm to data subjects.[129] For some, regulations were seen as being too generic, with some research exempt from consent requirements (e.g. cancer registries), although the rationale for such policies is not always transparent and/or consistent.[130]

The right to privacy as an individual-level right and the general encouragement of 'open data' appear to clash for some respondents, conceptually but also as a matter of scientific culture.[131] Regulation was seen as lagging behind practice, e.g. an open data approach is required to receive research funding, but the legal framework applicable to cloud computing is unclear, and issues such as potential for genetic discrimination or the consent of children (in the case of genome sequencing) are insufficiently or inadequately addressed. Finally, there is disagreement and potential inconsistency concerning the feasibility and desirability of anonymising genomic datasets and also the use of alternative mechanisms such as 'safe havens'.[132]

What examples are there of innovative initiatives that promote privacy while encouraging participation?

One suggestion, aside from more earnest public engagement, was to make use of 'dynamic consent', with some respondents in favour and at least one response strongly against.[133] As with other questions, the disagreement might, however, be partly due to different interpretations of what this approach would entail. Some warned that technological fixes for privacy concerns are unlikely to be 'future-proof'.[134] One response suggested that, in particular, the 100,000 Genomes Project currently developed by Genomics England "could provide an excellent opportunity to inform this process."[135]

[128] Cf. Prof. S. Bird, p. 4.
[129] E.g.The Wellcome Trust, p. 12.
[130] PHG Foundation, p. 16.
[131] N. Levin, p. 6.
[132] UCL Centre for Health Informatics and Multiprofessional Education, p. 4.
[133] PHG, p. 17; UCL Centre for Health Informatics and Multiprofessional Education, p. 2; The Wellcome Trust, p. 11. Cf. The Mason Institute, p. 14.
[134] Royal Academy of Engineering, p. 3.
[135] The Wellcome Trust, p. 3.